A general theory of DNA-mediated and other valence-limited colloidal interactions

Patrick Varilly, Stefano Angioletti-Uberti, Bortolo M. Mognetti and Daan Frenkel[1]

Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, UK

(Dated: 6 March 2013)

We present a general theory for predicting the interaction potentials between DNA-coated colloids, and more broadly, any particles that interact via valence-limited ligand-receptor binding. Our theory correctly incorporates the configurational and combinatorial entropic factors that play a key role in valence-limited interactions. By rigorously enforcing self-consistency, it achieves near-quantitative accuracy with respect to detailed Monte Carlo calculations. With suitable approximations and in particular geometries, our theory reduces to previous successful treatments, which are now united in a common and extensible framework. We expect our tools to be useful to other researchers investigating ligand-mediated interactions. A complete and well-documented Python implementation is freely available at `http://github.com/patvarilly/DNACC`.

I. INTRODUCTION

The theoretical study of the statistical behavior of soft matter systems generally begins with a description of the interaction between their constituent particles. One exciting system that has emerged in recent years is that of DNA-coated colloids (DNACCs), where interactions between colloids are driven by partial hybridization of complementary ssDNA strands grafted on their surface[1-4]. To date, DNACCs have been used mainly for building highly sensitive biomedical probes[5] and to assemble binary nanoparticle crystals of numerous morphologies[6-9].

The aim of this paper is to present a general theory of DNACC interactions that can be applied to any DNACC system, using only mild assumptions that are valid for common DNACCs. A key feature of our approach is that once the binding statistics of an arbitrary pair of DNA strands is known, all combinatorial and competitive effects can be computed separately. Disentangling these two aspects is what gives our theory its generality. Our efforts build on past theoretical attempts to understand DNACC experiments in terms of the statistical mechanics of bond formation[3,10-18], which have all partially succeeded in modeling various experimental setups, differing for instance in coating densities, nucleotide sequences, strand lengths, stiffness and colloid sizes. To date, however, a rigorous and general solution to the problem has not been presented[17-19].

A salient feature of DNACCs is that their interaction is mediated by surface-grafted ligands, each of which can form a single bond at a time. Owing to this restriction, we say that such interactions are "valence-limited"[20]. While we phrase our theory in the language of DNACCs, the treatment is equally applicable to any valence-limited interaction, including multi-valent particles for drug delivery, viruses and cells[21-24].

The rest of the paper is structured as follows. In Section II, we present a statistical mechanical derivation of our theory in its most general form, followed by a specialization to DNACCs and a computationally simpler mean-field approximation. In Section III, we compare the results of our theory with those from more accurate and computationally expensive Monte Carlo simulations. We have chosen as examples specific DNACC systems that have been presented previously[15-17,25], as well as two novel ones, to show that they can all be described using a single, unifying framework. In Section IV, we summarize our approach and reproduce the essential equations for reference purposes. Finally, in Section V, we discuss possible applications of our theory, both for DNACCs and for more general valence-limited interactions.

In Appendix A, we discuss an important technical aspect of our theory, namely, how to calculate the entropic cost of binding for two general polymeric ligands. We also summarize the results for short, stiff rods (which model short dsDNA tethers common in micron-sized DNACCs[3,4,14]) and ideal chains (as used in Refs. 17 and 18). Appendix B explains how our theory is connected to the Local Chemical Equilibrium treatment recently introduced by Rogers and Crocker[19] (itself an improvement of the work of Biancaniello, Kim and Crocker in Ref. 3). We show how a physically motivated and simple change to the equations used by these authors corrects their theory's principal shortcoming, and that the corrected version reduces to a specialization of our theory. Finally, Appendix C presents an alternate derivation of the mean-field approximation to our theory using a saddle-point analysis, making explicit the connection between our theory and earlier successful but more specialized approaches[16,25].

Although this paper explains our approach in full detail, we emphasize that the final results are simple and can be used widely by other researchers to calculate pair potentials and binding configurations of general DNACCs. To this end, we have made freely available a complete and well-documented Python implementation of the theory, which can be found at `http://github.com/patvarilly/DNACC`.

II. THEORETICAL MODEL

Under experimentally relevant conditions, DNA-coated colloids interact through entropic repulsion and hybridization-mediated attraction[3,12,15]. Whereas the repulsion is simple to estimate for common geometries, as shown in Appendix A, calculating the attraction is more complicated. The key subtlety is that the attraction is mediated by valence-limited binding, i.e., each DNA strand can bind at most one partner at a time[26]. In this section, we first present a general treatment of this attraction for generic valence-limited ligands. Then, we specialize our treatment to DNACCs. Finally, we construct a mean-field approximation that is suitable for geometries with translational symmetry.

A. Statistical Mechanics of valence-limited interactions

Consider a system of N generalized ligands, depicted in Figure 1(a). At any given time, an individual ligand i can bind at most one other ligand j, with a resulting free energy change ΔG_{ij}. If i and j cannot bind, either for geometric or chemical reasons, we set ΔG_{ij} to infinity. Since i cannot bind to itself, we also set ΔG_{ii} to infinity for all i. The set of all ΔG_{ij} values completely characterizes the relevant chemical detail of the system. We explicitly assume that ΔG_{ij} is independent of the presence of other ligands, which is accurate when the density of ligands is low enough that two ligands not bound to each other can be thought of as not interacting at all. This condition holds for the grafting densities and salt concentrations typically used in DNACC experiments with micron-sized colloids[3,12,14,27].

Let ϕ be a binding configuration of the system, i.e. a list of bound ligand pairs (i, j). Taking the state where no bonds are formed as a reference state, the free energy of ϕ is the sum of the binding free energies of each bond. Summing over all ϕ, we obtain the exact partition function of the system, Z, relative to that of the reference state, Z_0:

$$\frac{Z}{Z_0} = \sum_{\{\phi\}} \prod_{(i,j) \in \phi} e^{-\beta \Delta G_{ij}}. \tag{1}$$

For a particular ligand i, the terms in the sum above can be classified according to the binding partner of i, if any. Let Z_{-i} be the partition function for a system where i is unbound, and let $Z_{-i,-j}$ likewise be the partition function for a system where both i and j are unbound. The sum of the terms in Eq. (1) where i is unbound is Z_{-i}/Z_0, whereas the terms where i is bound to a ligand j out of the remaining $N-1$ ligands is $\exp(-\beta \Delta G_{ij}) Z_{-i,-j}/Z_0$. With these definitions, it follows from Eq. (1) that

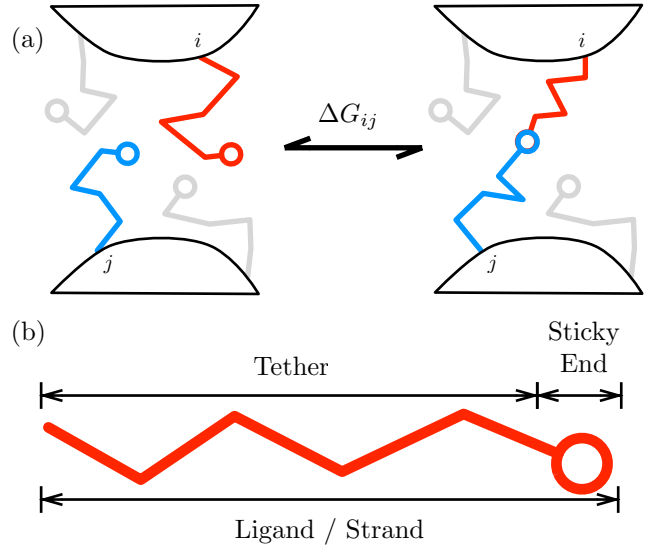$$1 = \frac{Z_{-i}}{Z} + \sum_j e^{-\beta \Delta G_{ij}} \frac{Z_{-i,-j}}{Z}, \qquad \text{for all } i, \tag{2}$$

where the sum is over the $N-1$ ligands j distinct from i. Let p_i be the probability that ligand i is unbound:

$$p_i = Z_{-i}/Z. \tag{3}$$

In Eq. (2), the factor $Z_{-i,-j}/Z$ is the probability that both i and j are unbound. Noting that $Z_{-i,-j}/Z_{-j}$ is the conditional probability for i to be unbound given that j is unbound, we can rewrite $Z_{-i,-j}/Z$ as

$$\frac{Z_{-i,-j}}{Z} = \frac{Z_{-i,-j}}{Z_{-j}} \frac{Z_{-j}}{Z} \approx \frac{Z_{-i}}{Z} \frac{Z_{-j}}{Z} = p_i p_j. \tag{4}$$

Above, we have approximated the binding of i and j as uncorrelated events. This approximation is accurate if both i and j have many possible binding partners with comparable binding free energies.

Using Eqs. (2)–(4), we obtain the following set of N self-consistent equations for the quantities $\{p_i\}$:

$$1 = p_i + \sum_j e^{-\beta \Delta G_{ij}} p_i p_j, \qquad \text{for all } i. \tag{5}$$

When rewritten in the following convenient form, these equations can be solved, for example, by self-consistent iteration:

$$p_i = \frac{1}{1 + \sum_j e^{-\beta \Delta G_{ij}} p_j}, \qquad \text{for all } i. \tag{6}$$

The probability p_{ij} that ligands i and j are bound is then given by

$$p_{ij} = e^{-\beta \Delta G_{ij}} p_i p_j. \tag{7}$$

FIG. 1. (a) Schematic of a DNACC system. Any pair of ligands, i and j, can form a bond at free energy cost ΔG_{ij}, which is approximately independent of the presence of the other ligands. (b) Anatomy of a typical DNACC ligand, consisting of a long tether capped by a point-like sticky end. Throughout, we use the terms ligand and strand interchangeably.

Eqs. (6) and (7) are a principal result of this paper. It can be shown that these results are a single-particle analog of the more familiar chemical equilibrium conditions between the concentrations of N species of particles, with binding constants proportional to $\exp(-\beta\Delta G_{ij})$.

Eq. (6) can be recast as a minimization problem, which leads to a more efficient method of solution than fixed-point iteration. A similar observation has recently been made when solving the equations that arise in umbrella sampling free energy calculations (WHAM/MBAR), which have a similar structure[28,29]. Additionally, this form allows us to prove that Eq. (6) has exactly one solution, which satisfies $0 < p_i \leq 1$ for all i by construction. We define $f(\{p_i\})$, the function to be minimized, as follows:

$$f(\{p_i\}) = \sum_i [p_i - \ln p_i] + \frac{1}{2}\sum_{i,j} p_i k_{ij} p_j, \qquad (8)$$

where $k_{ij} = \exp(-\beta\Delta G_{ij})$. The stationary points of f are given by the solutions to the following system of equations,

$$0 = \frac{\partial f}{\partial p_i} = 1 - \frac{1}{p_i} + \sum_j k_{ij} p_j, \qquad (9)$$

which are equivalent to Eq. (6). We now demonstrate that *all* stationary points of f are local minima, implying there is a single stationary point that is the global minimum. To this end, we prove that the Hessian of f at a stationary point, $\partial^2 f/\partial p_i\,\partial p_j$, is positive definite, implying that the stationary point is a local minimum, by showing that the quantity $\sum_{ij} v_i(\partial^2 f/\partial p_i\,\partial p_j)v_j$ is positive for all nonzero N-component vectors $\{v_i\}$. Indeed,

$$\sum_{ij} v_i \frac{\partial^2 f}{\partial p_i\,\partial p_j} v_j = \sum_{ij} v_i \left[\frac{1}{p_i^2}\delta_{ij} + k_{ij}\right] v_j, \qquad (10)$$

$$= \sum_{ij} \frac{v_i}{p_i}\left[\delta_{ij} + p_i k_{ij} p_j\right]\frac{v_j}{p_j}. \qquad (11)$$

The non-negative quantities in square brackets are the elements of a strictly diagonally dominant, symmetric matrix with positive diagonal elements, which is thus positive definite[30]. Indeed, the ith element of its diagonal is equal to $1 + p_i^2 k_{ii} \geq 1$ (we allow nonzero k_{ii} here, in anticipation of the mean-field self-consistent theory of Eq. (21) below), while the sum of the off-diagonal terms in the ith row satisfies

$$\sum_j p_i k_{ij} p_j - p_i^2 k_{ii} \leq \sum_j p_i k_{ij} p_j = 1 - p_i < 1, \qquad (12)$$

where the equality and the strict inequality both follow from Eq. (6) at the stationary point. Hence, the right-hand side of Eq. (11) is positive for every nontrivial choice of $\{v_i\}$, so the Hessian of f is positive definite, as claimed.

Having calculated the probabilities of each bond forming (Eq. (7)), we now calculate the resulting attraction free energy, F_{att}, given by

$$\beta F_{\text{att}} = -\ln(Z/Z_0). \qquad (13)$$

This quantity is most easily estimated using thermodynamic integration. One possible integration path is to offset all interaction energies by an amount λ/β, i.e.

$$\beta\Delta G_{ij} \mapsto \beta\Delta G_{ij} + \lambda. \qquad (14)$$

Let $Z(\lambda)$ be the partition function after this replacement. When λ is infinite, $Z(\lambda) = Z_0$, so

$$\beta F_{\text{att}} = \int_\infty^0 \mathrm{d}\lambda\, \frac{\partial[-\ln(Z(\lambda)/Z_0)]}{\partial\lambda} = -\int_0^\infty \mathrm{d}\lambda \sum_{i<j} p_{ij}(\lambda), \qquad (15)$$

where $p_{ij}(\lambda)$ is given by Eqs. (6) and (7) after effecting the substitution of Eq. (14).

Once the free energy is known at one temperature, a more physical integration path can be used to obtain it at any other temperature. To proceed, we denote the dependence of bond energies on β explicitly with the notation $\Delta G_{ij}(\beta)$, and similarly for $p_{ij}(\beta)$ and $F_{\text{att}}(\beta)$. We obtain

$$\beta_1 F_{\text{att}}(\beta_1) - \beta_0 F_{\text{att}}(\beta_0) = \int_{\beta_0}^{\beta_1} \mathrm{d}\beta \sum_{i<j} \frac{\partial\beta\Delta G_{ij}(\beta)}{\partial\beta} p_{ij}(\beta). \qquad (16)$$

In calculating the integral in Eq. (16) from β_0 to β_1, the free energy of binding at all intermediate temperatures is obtained as well.

Once F_{att} is known, the free energy of entropic repulsion, F_{rep}, must be added to obtain a full interaction free energy (see Appendix A for details). Additionally, it is usually convenient to offset all free energies to yield zero total energy when the colloids are an infinite distance apart, which we do in Section III.

In the limit of weak binding, when all the factors $\exp(-\beta\Delta G_{ij})$ are very small, the above results simplify significantly, as all the complications of binding competition disappear. Discarding all terms of order $\exp(-2\beta\Delta G_{ij})$ and higher, we obtain

$$p_i \approx 1 - \sum_j e^{-\beta\Delta G_{ij}}, \qquad (17)$$

$$p_{ij} \approx \exp(-\beta\Delta G_{ij}), \qquad (18)$$

$$\beta f_{\text{att}} \approx -\sum_{i<j} p_{ij}. \qquad (19)$$

In particular, the last relation can be used to sidestep the thermodynamic integration in Eq. (15), and is related to a similar "Poisson" approximation used in the literature[3,10,13,17,19]. However, we have previously pointed out that under realistic experimental conditions, this weak-binding approximation is rather poor[18], so its utility for experiments is limited.

B. Specializing to DNA-coated colloids

For a large class of ligands, binding occurs only in a small portion of the ligand. In DNACCs in particular,

the DNA strands that are used consist of a long, inert tether capped by a reactive, short sticky end, shown in Figure 1(b). The thermodynamics of association of DNA oligomers is well understood[31], so for a pair of strands, i and j, we can predict from their sequence the value of the hybridization free energy of their sticky ends in solution, ΔG_{ij}^0. Separating out this contribution to ΔG_{ij}, we have

$$\Delta G_{ij} = \Delta G_{ij}^0 + \Delta G_{ij}^{(\mathrm{cnf})}, \tag{20}$$

where $\Delta G_{ij}^{(\mathrm{cnf})}$ is the free energy penalty that arises from the reduced configurational space available to two bound ligands with respect to two free ligands[12,15]. This quantity depends only on the polymer statistics of the tethers and the way in which they are attached to the colloids (e.g., tethered on the surface vs. mobile), but not on the sticky ends, and in many cases is a completely entropic effect. In Appendix A, we list explicit expressions for $\Delta G_{ij}^{(\mathrm{cnf})}$ applicable to short, rigid dsDNA tethers grafted on plates. We also show how to use common polymer statistics techniques to estimate $\Delta G_{ij}^{(\mathrm{cnf})}$ to arbitrary accuracy for more complicated tethers, such as long, flexible ssDNAs modeled as ideal chains[3,17].

C. Mean field theory for DNA-coated plate interactions

In some regimes, notably micron-sized colloids with short DNA tethers, the range of the binding interactions is small compared to the radii of curvature of all the colloidal surfaces involved. In such cases, we can use the Derjaguin approximation[32] to calculate particle interaction energies, knowing only the interaction energy density of a pair of uniformly coated parallel plates as a function of plate separation. For this translationally invariant geometry, a substantial simplification is possible if the ligands are treated at a mean field level rather than individually.

The simplified system consists of two parallel plates separated by a distance h. We consider M different types of ligands, labeled by letters a and b below. Two ligand types can differ in their sticky end sequences, tether types or in the plate on which they are grafted. Corresponding to these M types, there are $M(M+1)/2$ potentially different solution hybridization free energies, denoted by ΔG_{ab}^0. Let σ_a be the grafting density of a-type tethers, which has units of tethers per unit area. Since we approximated the system as uniform, all a-type tethers have the same probability of being unbound, denoted by p_a. By approximating the sum in Eq. (6) by a continuous integral, we obtain a mean-field, self-consistent set of equations

$$p_a = \frac{1}{1 + \sum_b K_{ab}\sigma_b p_b}, \tag{21}$$

where

$$K_{ab} = e^{-\beta\Delta G_{ab}^0} \int \mathrm{d}^2\mathbf{r}_b \, \exp[-\beta\Delta G_{ab}^{(\mathrm{cnf})}]. \tag{22}$$

In the last equation, $\Delta G_{ab}^{(\mathrm{cnf})}$ is the configurational cost for a fixed a-type ligand to bind a b-type tether grafted at $\mathbf{r}_b$. The factor K_{ab} has units of area, and can be interpreted as an average Boltzmann factor for tether binding times the area on plate B within reach of a fixed a-type tether. A relation analogous to Eq. (7) yields the density of bonds between a-type and b-type tethers, σ_{ab}:

$$\sigma_{ab} = \sigma_a p_a K_{ab} p_b \sigma_b. \tag{23}$$

Similarly, using a relation analogous to Eq. (15), we can use thermodynamic integration to estimate the free energy of attraction per unit area, $f_{\mathrm{att}}(h)$, at fixed plate separation:

$$\beta f_{\mathrm{att}}(h) = -\int_0^\infty \mathrm{d}\lambda \sum_{a \leq b} \sigma_{ab}(\lambda), \tag{24}$$

where λ is the integration parameter defined in Eq. (14).

Eqs. (21)–(24) are a second principal result of this paper. They constitute a self-consistent mean field theory for the DNA-mediated attraction of uniformly coated parallel plates. In Appendix C, we show that, when specialized to only one or two kinds of bonds, these equations reduce to the earlier self-consistent mean field theories of Refs. 16 and 25. In these particular cases, we also show how Eqs. (21)–(24) may be derived from a saddle-point approximation of a mean-field partition function.

For the case of short, rigid dsDNA tethers, the factors K_{ab} have particularly simple expressions, derived in Appendix B. In particular, let tethers of type a and b have lengths L_a and L_b. Without loss of generality, let $L_a < L_b$. When a and b are on the same plate,

$$K_{ab} = \frac{e^{-\beta\Delta G_{ab}^0}}{\rho_0} \times \begin{cases} 1/L_b, & L_b < h; \\ 1/h, & \text{otherwise.} \end{cases} \tag{25a}$$

whereas when a and b are on opposite plates,

$$K_{ab} = \frac{e^{-\beta\Delta G_{ab}^0}}{\rho_0} \times \begin{cases} 0, & L_a + L_b < h; \\ \frac{L_a+L_b-h}{L_a L_b}, & L_b < h < L_a + L_b; \\ 1/h, & \text{otherwise.} \end{cases} \tag{25b}$$

When the tethers are modeled as freely-jointed chains of segments of length ℓ, a choice that has been suggested for modeling ssDNA tethers[17], the factors K_{ab} can be approximated with the following expressions, derived in Appendix A 2. Let N_a and N_b be the number of segments in tethers of type a and b. When a and b are on the same plate,

$$K_{ab} \approx \frac{e^{-\beta\Delta G_{ab}^0}}{\rho_0}$$
$$\times \sqrt{\frac{3}{\pi(N_a+N_b)\ell^2}} \, \mathrm{erf}\left(\sqrt{\frac{3h^2}{(N_a+N_b)\ell^2}}\right)$$
$$\bigg/ \mathrm{erf}\left(\sqrt{\frac{3h^2}{2N_a\ell^2}}\right) \mathrm{erf}\left(\sqrt{\frac{3h^2}{2N_b\ell^2}}\right), \tag{26a}$$

whereas when a and b are on opposite plates,

$$K_{ab} \approx \frac{e^{-\beta \Delta G_{ab}^0}}{\rho_0} \exp\left[-\frac{3h^2}{4\ell^2(N_a + N_b)}\right]$$
$$\times \sqrt{\frac{12}{\pi(N_a + N_b)\ell^2}} \, \mathrm{erf}\left(\sqrt{\frac{3h^2}{4(N_a + N_b)\ell^2}}\right)$$
$$\Big/ \mathrm{erf}\left(\sqrt{\frac{3h^2}{2N_a \ell^2}}\right) \mathrm{erf}\left(\sqrt{\frac{3h^2}{2N_b \ell^2}}\right). \quad (26b)$$

In both of these equations, the error function, $\mathrm{erf}(x)$, is defined as

$$\mathrm{erf}(x) = \frac{2}{\pi} \int_0^x \mathrm{d}t \, e^{-t^2}. \quad (27)$$

As before, when all the factors $K_{ab}\sigma_b$ and $\sigma_a K_{ab}$ are very small, one can implement a weak-binding approximation to sidestep the thermodynamic integration in Eq. (15), which yields

$$p_a \approx 1 - \sum_b K_{ab}\sigma_b, \quad (28)$$

$$\sigma_{ab} \approx \sigma_a K_{ab}\sigma_b, \quad (29)$$

$$\beta f_{\mathrm{att}} \approx -\sum_{a \leq b} \sigma_{ab}. \quad (30)$$

However, as pointed out previously, the utility of this approximation under realistic experimental conditions is limited.

Frequently, DNACCs are covered uniformly at a fixed density σ with a single type of strand, and only the sticky ends vary from one colloid to the next. In this case, we can obtain an explicit mean-field expression for βf_{att} between a plate grafted with a-type strands and one grafted with b-type strands. Defining J as the dimensionless factor $2K_{ab}\sigma$, we obtain

$$\sigma_{ab} = \frac{\sigma}{J}\left[(1+J) - \sqrt{1+2J}\right], \quad (31)$$

so,

$$\beta f_{\mathrm{att}} = \frac{\sigma}{J}\Big[-(1+J) + \sqrt{1+2J}$$
$$+ 2J\ln\Big\{\frac{1}{2}(1 + \sqrt{1+2J})\Big\}\Big]. \quad (32)$$

This expression makes explicit the breakdown of the weak-binding approximation (in this case, $\beta f_{\mathrm{att}} \approx -\sigma_{ab}$) when $J \gtrsim 1$. Unfortunately, even for the simple case of rod-like DNA strands, the above expression does not yield a simple analytic result for the pair potential between to DNA-coated spheres via the Derjaguin approximation. Since this difficulty only increases when we allow for uneven coverages and multiple potential interactions, we believe that our numerical thermodynamic integration procedure is justified.

III. RESULTS AND DISCUSSION

In this section, we compare the results of detailed Monte Carlo simulations (MC) with those derived from our theory, described above. We refer to the results of the theory where each strand is modeled individually (Eqs. (6), (7) and (15)) as the SCT results, and those of the form where the strands are modeled at a mean field level (Eqs. (21)–(24)) as the SCTMF results[33].

A. Colloids coated with short dsDNA strands

The simplest possible DNACC is a flat plate that is uniformly covered by a single type of DNA strand. Interaction energies obtained from plate geometries can be used to approximate interaction energies in more complex geometries by use of the Derjaguin approximation. Although, our theory can be used in arbitrary geometries, in this section we focus on plates for clarity.

The interaction free energy per unit area of two such DNACCs has been calculated previously from Monte Carlo simulations[15,16]. In this case, the tethers were modeled as short, stiff rods of length $L = 20\,\mathrm{nm}$, grafted at a density of $1/(20\,\mathrm{nm})^2$. Figure 2 shows the fraction of bonds formed between the two plates when separated by a distance $h = L$, as a function of the sticky end solution hybridization free energy, ΔG^0. The results are calculated both from Monte Carlo simulations and from our theory, using an identical realization of the tether grafting positions on a periodic $200\,\mathrm{nm} \times 200\,\mathrm{nm}$ plate. The figure inset shows the free energy per unit area of these plates, as a function of plate separation, for three different values of ΔG^0.

For the grafting density used in Figure 2, each tether typically has around 9 possible binding partners, so the mean-field approximation is excellent. Small discrepancies arise only at low temperature and large plate separation, where the number of possible binding partners is low and Poisson fluctuations in local coverage density are relevant. Such fluctuations are naturally accounted for in the explicit self-consistent theory, whose predictions are thus nearly indistinguishable from MC simulations in all cases.

Figure 3 shows the analogous results for a pair of DNACCs coated with two kinds of tethers which has been described previously[25]. We label the tethers on one plate by a and b, and those on the other plate by a' and b'. Each tether type is grafted at a density of $1/(20\,\mathrm{nm})^2$. The sticky ends of a and a' hybridize strongly, with free energy ΔG_a^0, forming bridges between the two plates. Conversely, the sticky ends of a and b, as well as those of a' and b', hybridize weakly, with free energy $\Delta G_a^0 + 5\,k_{\mathrm{B}}T$, forming loops within each plate. A competition thus arises between strong bridges and weak loops, which has been shown to result in interaction potentials with an unusual temperature dependence that should significantly enhance the ability of these DNACCs
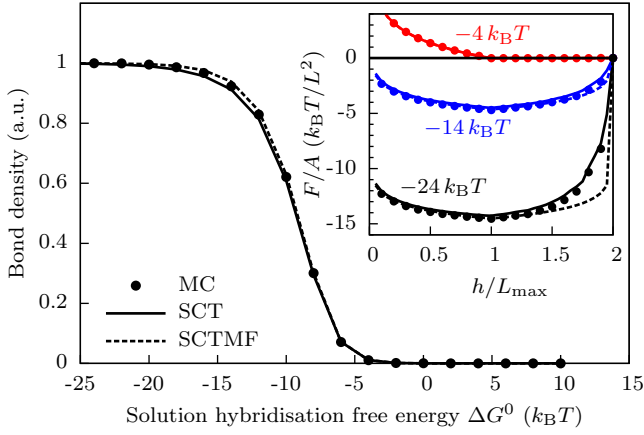
FIG. 2. Density of bonds formed between two flat plates coated uniformly with short, rigid dsDNA tethers of length $L = 20\,\text{nm}$, at plate separation $h = L$, as a function of the sticky end solution hybridization free energy, ΔG^0. Inset: potential of mean force per unit area as a function of plate separation, for three different values of ΔG^0.
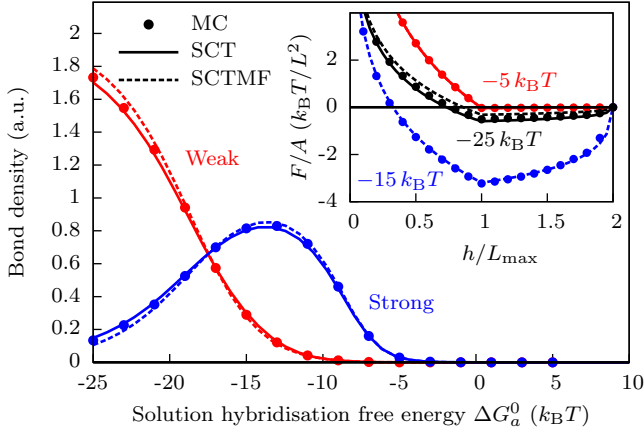

FIG. 3. Density of strong and weak bonds formed between plates with competing-linkages, at plate separation $20\,\text{nm}$. Strong a-a' inter-plate bonds (bridges) have sticky end hybridization free energy ΔG_a^0, and weak a-b and a'-b' intra-plate bonds (loops) have sticky end hybridization free energy $\Delta G_a^0 + 5\,k_\text{B}T$. See text for details. Inset: potential of mean force per unit area as a function of plate separation, for $\Delta G_a^0 = -5\,k_\text{B}T$ (upper), $-15\,k_\text{B}T$ (bottom) and $-25\,k_\text{B}T$ (middle). When the sticky end binding is strongest, the plates barely interact.

to form binary crystals[25]. All the features of these interactions are reproduced with quantitative accuracy by our self-consistent theory.

The key advantage of the present theory is its generality and comparative simplicity. With little additional effort, we can consider systems with many different kinds of grafted DNAs, each with different entropic costs for binding, and with different and potentially nonuniform
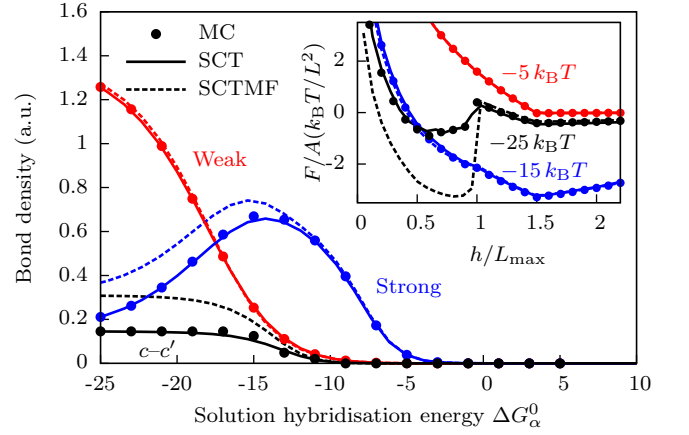

FIG. 4. Density of strong, weak and c–c' bonds formed in the third system, at plate separation $20\,\text{nm}$ (see text for details). Inset: potential of mean force per unit area as a function of plate separation, for three values of ΔG_a^0.

coating densities. In the previous two cases, the resulting SCTMF theory is equivalent to earlier self-consistent mean-field theories that were less general and that were derived more heuristically (Appendix C). In our third and fourth examples, we go beyond what has been done previously to illustrate the generality of our approach.

For our third example, we generalize the setup of Figure 3 to include a third type of tether on each plate, respectively c and c'. The sticky ends of c and c' hybridize only with each other and with the same strength as the a and a' sticky ends. Furthermore, we set the lengths of the a, b, a' and b' tethers to $30\,\text{nm}$ and those of the c and c' tethers to $10\,\text{nm}$. Finally, we set the grafting density of the c and c' tethers to $1/(30\,\text{nm})^2$. Thus, our third system features three different types of rods, of different lengths, with different coverages and with a nontrivial interaction matrix. Figure 4 summarizes the behavior of this system. Our choices result in the interaction energy per unit area having up to two different minima, depending on the specific temperature (binding strength). The example also illustrates one way in which the mean-field approach fails, and how the full theory behaves for such cases. The large discrepancies in the number of c–c' bonds formed (black solid and dashed curves in Figure 4), and the resulting discrepancies in the potentials of mean force, result from the low coverage density of these tethers. The typical separation between grafting points that result from Poisson statistics is far larger than the length of the tethers, but these inhomogeneities are not captured by a mean field approach. Importantly, since the full theory retains detailed information on the grafting points of c and c' tethers, the MC results are again recovered to quantitative accuracy.

Our fourth and final example in this section is a system of two parallel plates, each with a $100\,\text{nm}$-radius circular patch of dsDNA strands, distributed uniformly within the patch at a coverage density $1/(20\,\text{nm})^2$. As
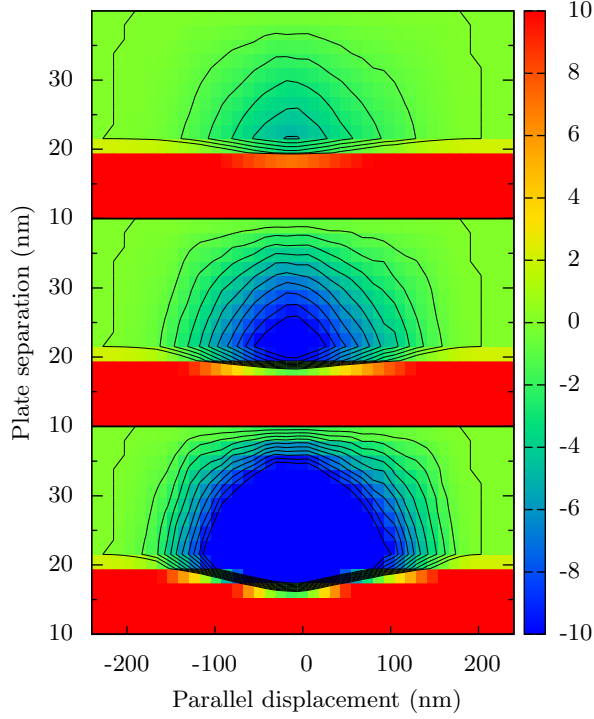
FIG. 5. Potential of mean force for two parallel plates coated with circular patches (radius $100\,\text{nm}$) of short dsDNA strands with complementary sticky ends. The coating density is $1/(20\,\text{nm})^2$. The sticky ends have solution hybridization free energy of $-6\,k_\text{B}T$ (top), $-7\,k_\text{B}T$ (middle) and $-8\,k_\text{B}T$ (bottom). Contours are drawn with $1\,k_\text{B}T$ spacing.

above, the strands are modeled as $20\,\text{nm}$ rigid rods. The only hybridization allowed is between strands on opposite plates, and we consider solution hybridization energies of $-8\,k_\text{B}T$, $-7\,k_\text{B}T$ and $-6\,k_\text{B}T$. Figure 5 shows the potentials of mean force of these plates as a function of plate separation and lateral displacement (zero lateral displacement means that the centers of the circular patches line up vertically). Such detailed potentials could be used to inform parameter choices for coarse-grained models of patchy particles[34] when the patches are made from patterned DNA coatings. For nanoscopic colloids, we suggest that such patches could be realized by building the particles through DNA origami[35]. However, to our knowledge, such precise experimental control of coating patterns for microscopic colloids remains elusive, although recent experiments with triblock Janus particles are very encouraging[36].

B.   Colloids coated with long ssDNA tethers

To describe more flexible ssDNA tethers, we use the representation recently employed by Rogers and Crocker[17] to describe experimental measurements of
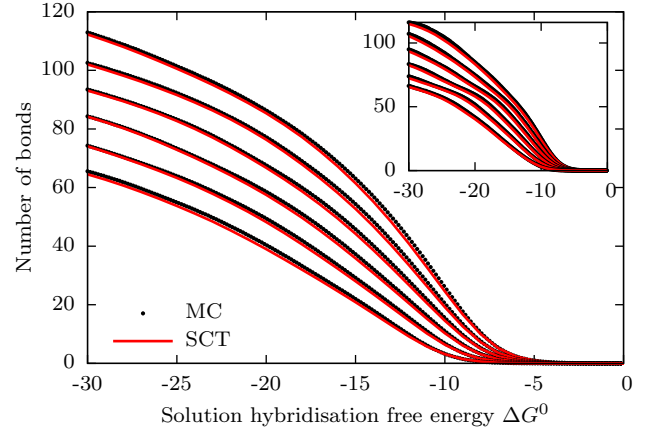

FIG. 6. Calculated number of bonds formed between the A- and B-type colloids described in Ref. 17, when separated by $5\,\text{nm}$ (top curve) to $30\,\text{nm}$ (bottom curve) in steps of $5\,\text{nm}$, averaged over 14 realizations of tether grafting points on the colloids. Inset: Same before averaging, i.e., for an individual realization of tether grafting points.

DNACC interaction potentials. They model ssDNA tethers of contour length $40\,\text{nm}$ as 8-segment freely jointed chains of Kuhn length $5\,\text{nm}$. The sticky ends that cap them are modeled as point-like, so that when two tethers bind, they form a 16-segment freely jointed chain. Appendix A 2 describes how to calculate $\Delta G_{ij}^{(\text{cnf})}$ for two such tethers i and j.

In Rogers and Crocker's experiment, there are two complementary types of sticky ends, A and B, whose sequences are predicted[37] by DINAmelt[38] to have binding enthalpy $\Delta H^0 = -47.8\,\text{kcal/mol}$ and binding entropy $\Delta S^0 = -139.6\,\text{cal/(mol·K)}$. At temperature T, the A-B binding free energy in solution is taken to be $\Delta G^0 = \Delta H^0 - T\Delta S^0$. Tethers with these sticky ends are coated on three types of spherical colloids, all of diameter $1.1\,\mu\text{m}$, A-type colloids are coated with 4800 ± 480 A-type tethers, B-type colloids are coated with 4200 ± 420 B-type tethers, and AB-type colloids are coated with 2400 ± 240 A-type and 3500 ± 350 B-type tethers. In all the following simulations, we have used the central value of all ranges.

We first address the interaction between A and B colloids. The coating densities used by Rogers and Crocker are sufficiently low that local grafting density fluctuations are relevant. We have thus prepared 14 independent configurations of uniformly coated A and B colloids. Figure 6 shows the number of bonds formed between two such colloids when separated by $5\,\text{nm}$ to $30\,\text{nm}$ versus sticky end binding strength ΔG^0, as calculated by explicit Monte Carlo simulation and the explicit SCT theory, averaged over the 14 independent coating realizations. The agreement is nearly quantitative, both when averaged over grafting realizations and independently for each such realization (inset).

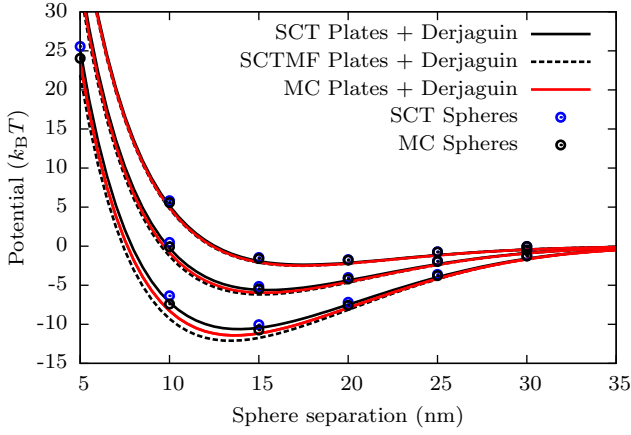By using Eq. (15), we can integrate the curves in Fig-

FIG. 7. Potential of mean force between A and B colloids of Ref. 17 at temperatures of 36.0°C, 33.0°C and 30.5°C (upper, middle and lower curves). See text for details of the various methods.
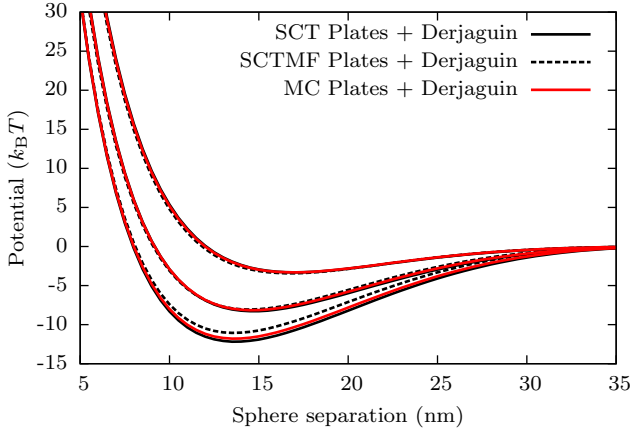

FIG. 8. Potential of mean force between two AB colloids of Ref. 17 at temperatures of 32.0°C, 27.0°C and 24.0°C (upper, middle and lower curves). See text for details.

ure 6 to obtain the mean attraction between the A and B colloids. Further accounting for the entropic repulsion caused by confining the tethers between the two colloids results in a potential of mean force for this system, shown as dots in Figure 7. In this Figure, we also show the result of calculating the colloid interaction potential by calculating the interaction energy per unit area between two similarly coated plates, and then applying the Derjaguin approximation. The interaction free energy between plates is calculated in three different ways: by explicit MC simulations, by using SCT, and by using its mean-field approximation, SCTMF. Our results here help establish the Derjaguin approximation as valid for ssDNA tethers modeled as freely-jointed chains. For completeness, we show similar results for the AB–AB system explored by Rogers and Crocker in Figure 8.

It should be noted that the simple freely-jointed chain

model of ssDNA used here and in Ref. 17 may not adequately capture the complications of the experimental system[18,19]. Our aim in this paper is limited to establishing the SCT as an accurate and fast alternative to detailed MC calculations, regardless of the underlying model used to calculate the configurational entropic factors associated with individual tethers binding.

IV. SUMMARY OF RESULTS

For ease of reference, in this section we summarize the main results of the paper. These results are implemented in the Python package DNACC, available at http://github.com/patvarilly/DNACC.

A. Self-consistent theory (SCT)

Our central result describes the interaction details of a DNACC system where each strand i is modeled explicitly. The algorithm is:

1. Using a suitable microscopic model, calculate the binding free energy ΔG_{ij} between every pair of strands i and j. When i equals j, or when i and j cannot bind, set ΔG_{ij} to infinity. This binding free energy is the sum of two contributions: the solution hybridization free energy of the sticky ends, ΔG_{ij}^0, and the entropic cost of binding of the two tethers that hold the sticky ends, $\Delta G_{ij}^{(\mathrm{cnf})}$. We provide expressions for $\Delta G_{ij}^{(\mathrm{cnf})}$ for rod-like and polymeric tethers in Appendix A.

2. Solve the self-consistent set of equations for each of the quantities p_i, i.e., the probability p_i that strand i is unbound:

$$p_i = \frac{1}{1 + \sum_j e^{-\beta \Delta G_{ij}} p_j}, \qquad \text{for all } i. \qquad (33)$$

It is straightforward to solve these equations using fixed-point iteration, i.e., making an initial guess for $\{p_i\}$, then repeatedly inserting the values of $\{p_i\}$ into the right-hand side of the above equations to obtain an improved estimate for $\{p_i\}$. A solution can also be obtained through minimization of a function with no other local minima or saddle points, as detailed in Section II A.

3. Calculate p_{ij}, the probability p_{ij} that strands i and j are bound:

$$p_{ij} = e^{-\beta \Delta G_{ij}} p_i p_j. \qquad (34)$$

The interaction free energy of a DNACC system can then be calculated through thermodynamic integration of the binding free energies ΔG_{ij}, as follows:

1. Calculate the the free energy of repulsion, $F_{\rm rep}$, equal to the free energy of the system when none of the strands can bind, i.e., when all the ΔG_{ij} are infinite. This is simply the sum of the free energies of repulsion of the individual strands. We assume this quantity can be calculated analytically or otherwise, as described in Appendix A.

2. Define a thermodynamic integration variable, λ, and denote by $p_{ij}(\lambda)$ the probability that strands i and j are bound when the all the binding free energies are respectively set to $\Delta G_{ij} + \lambda\, k_{\rm B}T$, calculated as above.

3. The free energy of the system is given by the following equation

$$F = F_{\rm rep} - k_{\rm B}T \int_0^\infty {\rm d}\lambda \sum_{i<j} p_{ij}(\lambda). \qquad (35)$$

B. Self-consistent mean-field theory (SCMFT)

When colloids are coated uniformly with a small number of different DNA species at high density, it is useful to model the colloid interaction in mean field. Assuming that the colloidal surfaces have radii of curvature significantly larger than the range of interaction of two strands, we may first calculate the interaction free energy of two parallel plates as a function of plate separation, h, then apply the Derjaguin approximation.

The details of the interaction between two parallel DNA-coated plates separated by a distance h are estimated as follows:

1. For each pair of DNA species, a and b, calculate a binding factor K_{ab} as follows:

$$K_{ab} = e^{-\beta \Delta G^0_{ab}} \int {\rm d}^2\mathbf{r}_b \, \exp[-\beta \Delta G^{\rm (cnf)}_{ab}], \qquad (36)$$

where $\Delta G^{\rm (cnf)}_{ab}$ is the configurational cost for a fixed a-type ligand to bind a b-type tether grafted at $\mathbf{r}_b$. If a DNA species is present in both parallel plates, treat the DNA on each plate as a separate species. Simple analytic expressions for K_{ab} for rod-like and polymeric tethers are given in Eqs. (25) and (26).

2. Solve the self-consistent set of equations for each of the quantities p_a, i.e., the probability that an a-type strand is unbound:

$$p_a = \frac{1}{1 + \sum_b K_{ab}\sigma_b p_b}, \qquad (37)$$

where σ_b is the number of b-type strands per unit area.

3. Calculate the number of a- and b-type bonds per unit are, σ_{ab}, as follows:

$$\sigma_{ab} = \sigma_a p_a K_{ab} p_b \sigma_b. \qquad (38)$$

The free energy per unit area between two DNA-coated parallel plates separated by a distance h is estimated as follows:

1. Calculate the free energy density of repulsion, $f_{\rm rep}(h)$, equal to the free energies of repulsion of the individual strand types, weighted by their coverage density (Eq. (A8)). Expressions for this quantity in the case of rod-like and polymeric ligands are given in Eqs. (A8), (A15) and (A33).

2. Define a thermodynamic integration variable, λ, and denote by $\sigma_{ab}(\lambda)$ the number of a–b bonds per unit area when the all the solution binding free energies are respectively set to $\Delta G^0_{ab} + \lambda\, k_{\rm B}T$, calculated as above.

3. The free energy density of the system is given by the following equation

$$f(h) = f_{\rm rep}(h) - k_{\rm B}T \int_0^\infty {\rm d}\lambda \sum_{a \le b} \sigma_{ab}(\lambda). \qquad (39)$$

Finally, to estimate the free energy of interaction between arbitrary colloids, use the Derjaguin approximation[32]. In the case of two spheres of radius R_1 and R_2, respectively, with a distance of closest approach h, this approximation yields:

$$F(h) \approx \frac{2\pi}{R_1^{-1} + R_2^{-1}} \int_h^\infty {\rm d}h' \, f(h'). \qquad (40)$$

In using the Derjaguin approximation, it is important to shift all free energies appropriately to ensure that $f(h)$ is zero at infinite h.

V. CONCLUSIONS AND OUTLOOK

The theory for DNACC interactions that we have presented, both in its explicit form (Eqs. (6), (7) and (15)) and its mean-field approximation (Eqs. (21)–(24)) is capable of describing many different DNACC setups, varying in grafting density, sticky end interactions and tether details. The theory is robust and correctly incorporates the statistical mechanics of valence-limited binding, which plays a key role in the behavior of DNACCs, and it does so at much lower cost than full MC simulations. Extending the theory to treat general ligands within our approximation is straightforward, as exemplified in Appendix A for freely-jointed-chains.

When comparing our full theory to MC data in the context of DNACCs, we have found quantitative agreement. The mean-field approximation, which substantially generalizes earlier approaches, is in quantitative to semi-quantitative agreement with the exact results. Significant discrepancies arise only when the grafting density fluctuations that are neglected in the mean-field approach become relevant.

Our theory accounts for all the complications arising from valence limitation and combinatorial entropy. Nevertheless, the binding free energy ΔG_{ij} of a pair of ligands i and j is taken as an external input, and the agreement of the theory's predictions with experimental results will depend significantly on the quality of the estimates for the values of ΔG_{ij}. Simple models for generating these estimates, such as those used in this paper, can inform experiments, though apparently, they are not yet sufficiently accurate for quantitative agreement with experiments, as we have discussed in Ref. 18.

We expect our theory to be a significant aid in the search for viable approaches to design crystal structures *a priori*. Presently, existing free energy methods can establish the stable crystal structure of a system given its interaction potential[39]. Conversely, robust methods are being developed for solving the inverse problem, determining the interaction potentials that might yield a particular target structure[40]. Designing colloids that realize such interaction potentials experimentally is another problem. In this regard, DNACCs, with their highly tunable interactions and vast parameter space (e.g., particle size, ligand structure and flexibility, coating density, nucleotide sequence length and hybridization strength), are a promising tool to build colloids with programmable interactions. For this reason, our theory can complement existing tools in crystal structure design by quickly suggesting experimentally realizable constructs that result in particular colloid–colloid interaction potentials. Whereas for simple potentials one can envision a manual search of parameter space guided by intuition and experience, it might also be possible to recast the problem in terms of a global minimization algorithm that optimizes the DNACC structures to obtain interaction potentials that are close to some target. It is precisely to help others easily use our tools in these and other ways that we have written a comprehensive and well documented series of Python modules that implement our theory, available under the GNU license at http://github.com/patvarilly/DNACC.

While we have developed our theory and the accompanying programs with an eye towards applying them to DNA-coated colloids, other interesting applications are possible. In particular, our theory can help model any system where attraction is dominated by linker-receptor-type bonding. These include patchy particles, where the patches are realized by a nonuniform ligand coat. More generally, many biological systems could be treated with our theory, for example, interactions between cells (or between cells and a substrate) that are mediated by membrane proteins. For these more general cases, care should be taken to include the effects of surface elasticity and how the entropic costs of ligand binding are, in turn, modified.

Appendix A: Entropic cost of ligand binding

In this appendix we describe how to calculate the free-energy change ΔG_{ij} of binding two ligands, i and j, when the binding interaction is physically limited to a small, reactive sticky end. We begin with a general treatment, and later specialize to the two important cases of short, stiff polymers and of freely-jointed chains. These two models have been used to describe the DNA strands on DNA-coated colloids when these are short, rigid dsDNA[12,15,16,25] and when they are long, flexible ssDNA[17,18]. These models may also be useful to describe ligand-mediated binding in important biological systems, such as cells, where adhesion is, in many cases, mediated by different glycoproteins coated on their surface membrane[41].

Let Q_i, Q_j and Q_{ij} be the partition functions of the two unbound ligands and the bound complex that they form. In terms of these quantities, ΔG_{ij} is given by

$$\exp(-\beta\Delta G_{ij}) = \frac{Q_{ij}}{Q_i Q_j}. \tag{A1}$$

As in the main text, we have assumed that the unbound ligands do not otherwise interact with each other. The Boltzmann factor $\exp(-\beta\Delta G_{ij})$ is generally smaller than that of the sticky ends in solution at standard conditions, $\exp(-\beta\Delta G_{ij}^0)$, owing to the entropic penalty of constraining the configurations available to ligands i and j when their reactive sticky ends are bound. This defines the configurational free energy cost of binding, $\Delta G_{ij}^{(\mathrm{cnf})}$, as follows:

$$\Delta G_{ij}^{(\mathrm{cnf})} = \Delta G_{ij} - \Delta G_{ij}^0. \tag{A2}$$

A realistic model that reproduces ΔG_{ij} for ligands such as DNA strands is expensive, as one would need to at least model each nucleotide of each strand explicitly[42]. It is computationally more convenient to separate the binding of the sticky ends from the configurational freedom of the rest of the ligand, and use a coarse-grained

representation of the sticky ends where ΔG_{ij}^0 is introduced as an explicit external parameter. Concretely, the internal partition functions of the free and bound sticky ends in solution, $\mathcal{Z}_i$, $\mathcal{Z}_j$ and $\mathcal{Z}_{ij}$, are related to the equilibrium constant of binding[14],

$$K = \frac{\exp(-\beta \Delta G_{ij}^0)}{\rho_0} = \frac{\mathcal{Z}_{ij}}{\mathcal{Z}_i \mathcal{Z}_j}, \tag{A3}$$

where ρ_0 is the standard concentration. Using this relation in Eq. (A1) to factorize the unknown dependence of Q_{ij}, Q_i and Q_j on the internal partition functions, we obtain

$$\frac{Q_{ij}}{Q_i Q_j} \approx \frac{\mathcal{Z}_{ij}}{\mathcal{Z}_i \mathcal{Z}_j} \frac{Q_{ij}^{(\mathrm{cnf})}}{Q_i^{(\mathrm{cnf})} Q_j^{(\mathrm{cnf})}}, \tag{A4}$$

where $Q_{ij}^{(\mathrm{cnf})}$, $Q_i^{(\mathrm{cnf})}$ and $Q_j^{(\mathrm{cnf})}$ are approximately independent of the chemical identity of the sticky end ("cnf" stands for "configurational"). It follows that

$$\exp[-\beta \Delta G_{ij}^{(\mathrm{cnf})}] \approx \frac{1}{\rho_0} \frac{Q_{ij}^{(\mathrm{cnf})}}{Q_i^{(\mathrm{cnf})} Q_j^{(\mathrm{cnf})}}. \tag{A5}$$

We stress that this factorization is an approximation that holds only when the binding of the sticky ends is only weakly affected by the presence of the rest of the ligand, as occurs when the ligand is an ideal polymer or when the binding is very strong. However, it is a useful approximation because once $\Delta G_{ij}^{(\mathrm{cnf})}$ has been measured or calculated, one can predict the values of ΔG_{ij} for families of ligands that differ only in their sticky ends, providing us with a powerful design tool.

We now specialize to the case of polymeric ligands with one end fixed and the other end sticky, such as polymers grafted at a surface. The main complication in treating polymers is dealing with excluded volumes. For instance, in DNACCs, the DNA strands cannot penetrate the colloids. In the discussion below, we denote by $\mathbf{r}_i$ the grafting point of polymer i. Given two polymers i and j, we denote by $\mathbf{r}_{ij}$ the vector $\mathbf{r}_j - \mathbf{r}_i$ with length r_{ij}. We denote by Ω the region of space that can be explored by the monomers of the polymers. Polymers restricted to Ω are deemed "confined", whereas those for which this restriction is absent are "unconfined".

We note that the free energy of repulsion due to a polymer i, denoted by $F_{\mathrm{rep},i}$, is captured by the dependence of $Q_i^{(\mathrm{cnf})}$ on the confinement region Ω. Indeed, letting $\hat{Q}_i^{(\mathrm{cnf})}$ be the partition function of polymer i in the absence of confinement, we have

$$e^{-\beta F_{\mathrm{rep},i}} = Q_i^{(\mathrm{cnf})}/\hat{Q}_i^{(\mathrm{cnf})}. \tag{A6}$$

From this expression follows the total free energy of repulsion,

$$F_{\mathrm{rep}} = \sum_i F_{\mathrm{rep},i}. \tag{A7}$$

Similarly, the mean-field free energy density of repulsion is given by

$$f_{\mathrm{rep}} = \sum_a \sigma_a F_{\mathrm{rep},a}, \tag{A8}$$

where σ_a is the coverage density of species a and $F_{\mathrm{rep},a}$ is the free energy of repulsion of any representative a-type strand.

1. Short, stiff polymers as ligands

A short, stiff polymer, such as a short piece of dsDNA, can be modeled as a rigid rod, as has been done previously in Refs. 15, 16, and 25. We gather the results here for completeness and to clarify the origin of a Jacobian prefactor absent in Ref. 15 but addressed in Ref. 16. For rigid rods, the configurational partition functions are:

$$Q_{ij}^{(\mathrm{cnf})} = \int_\Omega d\mathbf{r}\, \delta(|\mathbf{r} - \mathbf{r}_i| - L_i)\delta(|\mathbf{r} - \mathbf{r}_j| - L_j), \tag{A9a}$$

$$Q_i^{(\mathrm{cnf})} = \int_\Omega d\mathbf{r}\, \delta(|\mathbf{r} - \mathbf{r}_i| - L_i), \tag{A9b}$$

$$Q_j^{(\mathrm{cnf})} = \int_\Omega d\mathbf{r}\, \delta(|\mathbf{r} - \mathbf{r}_j| - L_j), \tag{A9c}$$

where L_i and L_j are the lengths of rods i and j.

The first integral is most easily evaluated in cylindrical coordinates, with the z axis parallel to $\mathbf{r}_{ij}$. Special care must be taken to isolate the Jacobian factor that arises when dealing with the product of the two δ-functions. After evaluating the integrals, we obtain,

$$Q_{ij}^{(\mathrm{cnf})} = \frac{\Omega_{ij}}{\sin(\gamma_i + \gamma_j)}, \tag{A10}$$

$$Q_i^{(\mathrm{cnf})} = \Omega_i, \tag{A11}$$

$$Q_j^{(\mathrm{cnf})} = \Omega_j. \tag{A12}$$

Here, Ω_{ij} is the portion within Ω of the circumference of the circle explored by the common endpoint of the hybridized rods, Ω_i and Ω_j are the areas within Ω of the sphere that the endpoints of the free rods can explore, and γ_i and γ_j are the angles that the hybridized rods make with $\mathbf{r}_{ij}$. The factor of $1/\sin(\gamma_i + \gamma_j)$ was obtained in Ref. 16 using two different representations of the δ-function, but the procedure here highlights the generality of the result.

When the colloids are plates, one at $z = 0$ and one at $z = h$, the region Ω is simply that where $0 \leq z \leq h$, and one can write an explicit expression for Ω_{ij}. For the case where the rods are grafted on opposite plates, such an expression is given in Appendix A of Ref. 16 (see also Figure 8 therein). When the rods are grafted to the same plate, Ω_{ij} is zero unless $|L_i - L_j| \leq r_{ij} \leq L_i + L_j$, in which case

$$\frac{\Omega_{ij}}{L_i \sin \gamma_i} = \begin{cases} \pi, & h \geq L_i \sin \gamma_i, \\ 2\arcsin[h/(L_i \sin \gamma_i)], & h < L_i \sin \gamma_i. \end{cases} \tag{A13}$$

The quantity Ω_i also has a compact expression,

$$\Omega_i = \begin{cases} 2\pi L_i^2, & h \geq L_i, \\ 2\pi L_i h, & 0 < h \leq L_i. \end{cases} \quad (A14)$$

The free energy of repulsion of a strand i follows immediately from Eq. (A6),

$$F_{\text{rep},i} = \begin{cases} -k_{\text{B}}T\ln(h/L_i), & 0 < h \leq L_i, \\ 0, & h \geq L_i. \end{cases} \quad (A15)$$

2. Long, flexible polymers as ligands

Following Refs. 17 and 18, we can model a polymer i as a freely-jointed chain (FJC) with N_i segments of length ℓ. Let $\mathbf{r}_{i,s}$ be the endpoint of segment s, with $1 \leq s \leq N$. For convenience, let $\mathbf{r}_{i,0} = \mathbf{r}_i$, and let $\mathbf{u}_{i,s} = \mathbf{r}_{i,s} - \mathbf{r}_{i,s-1}$. The chain segments are confined to the region Ω, but do not otherwise interact. It follows that

$$Q_i^{(\text{cnf})} = \int_\Omega \prod_{s=1}^{N_i} d\mathbf{u}_{i,s}\, \delta(|\mathbf{u}_{i,s}| - \ell), \quad (A16)$$

where, abusing notation somewhat, the integration domain confines the polymer to the region Ω. A completely analogous treatment yields $Q_j^{(\text{cnf})}$ for a FJC with N_j segments of length ℓ, grafted at $\mathbf{r}_j$.

One way of modeling the binding of i and j is to coalesce the endpoints of the two chains, so that $\mathbf{r}_{i,N_i} = \mathbf{r}_{j,N_j}$, resulting in a FJC with endpoints fixed at $\mathbf{r}_i$ and $\mathbf{r}_j$ and with $N_{\text{hyb}} = N_i + N_j$ segments $\{\mathbf{u}_s\}$, where $1 \leq s \leq N_{\text{hyb}}$. However, if the persistence length of the polymers is short, then the binding site may have a size comparable to the segment length ℓ. This is the case for ssDNA polymers with sticky ends that are 6 to 12 nucleotides long. Hence, we also consider a more general model for binding of i and j in which the final segment of the two polymers coalesces, so that $\mathbf{r}_{i,N_i} = \mathbf{r}_{j,N_j-1}$ and $\mathbf{r}_{i,N_i-1} = \mathbf{r}_{j,N_j}$, resulting in a FJC of $N_{\text{hyb}} = N_i + N_j - 1$ segments[43]. The partition function of the bound system is

$$Q_{ij}^{(\text{cnf})} = (4\pi\ell^2)^{N_i+N_j-N_{\text{hyb}}} \tilde{Q}_{ij}^{(\text{cnf})}, \quad (A17)$$

$$\tilde{Q}_{ij}^{(\text{cnf})} = \int_\Omega \Big[\prod_{s=1}^{N_{\text{hyb}}} d\mathbf{u}_s\, \delta(|\mathbf{u}_s| - \ell)\Big] \delta\Big(\mathbf{r}_{ij} - \sum_{s=1}^{N_{\text{hyb}}} \mathbf{u}_s\Big), \quad (A18)$$

where $\tilde{Q}_{ij}^{(\text{cnf})}$ is the partition function of a confined FJC with fixed end points, $\mathbf{r}_i$ and $\mathbf{r}_j$. The additional factor in $Q_{ij}^{(\text{cnf})}$ results from fusing the end segments of the individual polymers.

Although the integrals in Eqs. (A16) and (A18) cannot be evaluated analytically even in simple geometries, as was done in the previous section, one can calculate their value numerically using standard polymer techniques.

Let $Q_1(N) = (4\pi\ell)^N$ be the partition function of an unconfined N-segment FJC fixed at one endpoint, and let $Q_2(N, R)$ be the partition function of the same FJC with fixed end-to-end distance R. Their ratio yields the probability density for the end-to-end distance of the FJC to be R,

$$p(N, R) = \frac{Q_2(N, R)}{Q_1(N)}, \quad (A19)$$

which is known in closed form[44,45]. Using these partition functions, we can write $\tilde{Q}_{ij}^{(\text{cnf})}$ in the following convenient form:

$$\tilde{Q}_{ij}^{(\text{cnf})} = \frac{\tilde{Q}_{ij}^{(\text{cnf})}}{Q_2(N_{\text{hyb}}, r_{ij})} Q_2(N_{\text{hyb}}, r_{ij}). \quad (A20)$$

The ratio on the right-hand side of Eq. (A20) is the probability that a random walk from $\mathbf{r}_i$ to $\mathbf{r}_j$ lies wholly within Ω (i.e., does not penetrate the colloid). To compute it, we use Rosenbluth sampling[39], a common technique to generate a properly weighted ensemble of such FJCs. Specifically, to grow each segment s of one chain that starts at $\mathbf{r}_i$, we first generate K possible segments $\mathbf{u}_{s+1}^{(k)}$ ($1 \leq k \leq K$), sampled with probability density $p(N_{\text{hyb}} - (s+1), |\mathbf{r}_j - \mathbf{r}_{s+1}|)$ using rejection sampling[46]. We then pick at random one of the m_s segments in $\{\mathbf{u}_{s+1}^{(k)}\}$ wholly within Ω. The Rosenbluth weight of that chain is defined as

$$W = \prod_{s=1}^{N_{\text{hyb}}} \frac{m_s}{K}. \quad (A21)$$

The fraction of FJCs that lie wholly within Ω is then the average Rosenbluth factor,

$$\frac{\tilde{Q}_{ij}^{(\text{cnf})}}{Q_2(N_{\text{hyb}}, r_{ij})} = \langle W \rangle_{ij}, \quad (A22)$$

where the subscript implies that the average is computed as described above.

We can estimate $Q_i^{(\text{cnf})}$ and $Q_j^{(\text{cnf})}$ analogously, except that at each step in growing the chain, now of length N_i or N_j, the possible segments $\{\mathbf{u}_{s+1}^{(k)}\}$ are chosen uniformly on the surface of a sphere of radius ℓ, since the end of the chain is free. We obtain

$$\frac{Q_i^{(\text{cnf})}}{Q_1(N_i)} = \langle W \rangle_i, \quad (A23)$$

$$\frac{Q_j^{(\text{cnf})}}{Q_1(N_j)} = \langle W \rangle_j. \quad (A24)$$

Substituting Eqs. (A17)–(A24) into Eq. (A5), we obtain the following expression for the configurational entropy cost of binding i and j:

$$\exp(-\beta \Delta G_{ij}^{(\text{cnf})}) = \frac{p(N_{\text{hyb}}, r_{ij})}{\rho_0} \frac{\langle W \rangle_{ij}}{\langle W \rangle_i \langle W \rangle_j}. \quad (A25)$$
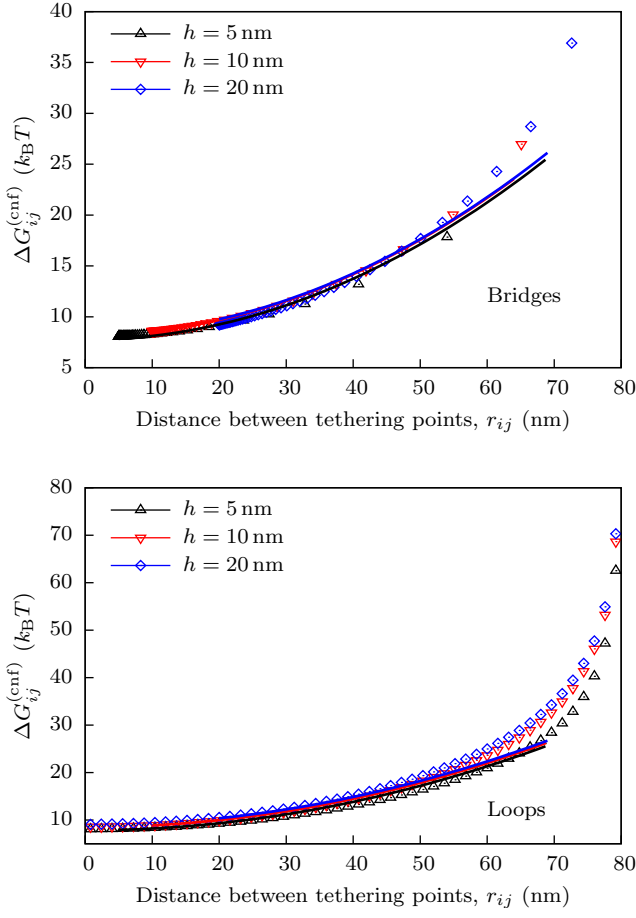
FIG. 9. $\Delta G_{ij}^{(\text{cnf})}$ for a plane geometry and the polymeric constructs of Sec. III B as obtained using MC simulations (symbols) and a harmonic spring approximation (Eqs. (A27), (A29) and (A30), full lines). The two FJCs are tethered either to different plates (above) or to the same plate (below).

In Figure 9 we show the configurational entropy cost of hybridizing two FJCs (made of 8 segments with $\ell = 5\,\text{nm}$) tethered to parallel plates. Interestingly, if $\Delta G_{ij}^{(\text{cnf})}$ is expressed in terms of the distance between tethering points, r_{ij}, and the plate separation, h, the dependence on the latter is marginal. This suggests that, for ideal chains, the cost of stretching a chain ($p(N_{\text{hyb}}, r_{ij})$ in Eq. (A25)) is more important than the excluded volume terms (W). Hence, we may approximate $\Delta G_{ij}^{(\text{cnf})}$ analytically by neglecting all chain degrees of freedom but the sticky end. In particular, using a Gaussian approximation for $p(N, r)$[45],

$$p(N, r) \approx \left(\frac{3}{2\pi N\ell^2} \right)^{3/2} \exp\left[\frac{-3r^2}{2N\ell^2} \right] \quad \text{(A26)}$$

we can model the ligand as a reactive point-like sticky end that is confined between the two walls, and is attached to the tethering point via a harmonic spring with spring constant $k = 3\,k_{\text{B}}T/N\ell^2$. The partition function of an unhybridized polymer is then given by

$$Q_i^{(\text{cnf})} \approx \int_\Omega \mathrm{d}\mathbf{r}\, p(N, |\mathbf{r} - \mathbf{r}_i|) = \frac{1}{2}\, \text{erf}\left[h\sqrt{\frac{3}{2N\ell^2}} \right], \quad \text{(A27)}$$

where the integration domain, Ω, is $[-\infty : \infty]^2 \times [0 : h]$, and the error function, $\text{erf}(x)$, is defined as in Eq. (27). Similarly, the partition function of the hybridized polymer is given

$$Q_{ij}^{(\text{cnf})} \approx \int_\Omega \mathrm{d}\mathbf{r}\, p(N, |\mathbf{r} - \mathbf{r}_i|) \cdot p(N, |\mathbf{r} - \mathbf{r}_j|). \quad \text{(A28)}$$

When i and j are tethered to opposite plates, so that the bound pair forms a bridge, we obtain

$$Q_{ij}^{(\text{bridge})} \approx \frac{1}{8}\left(\frac{3}{\pi N\ell^2} \right)^{3/2} \text{erf}\left(\frac{h}{2}\sqrt{\frac{3}{N\ell^2}} \right) \exp\left[-\frac{3r_{ij}^2}{4N\ell^2} \right],$$
$$\text{(A29)}$$

whereas when i and j are tethered to the same plate, so that the bound pair forms a loop, we obtain

$$Q_{ij}^{(\text{loop})} \approx \frac{1}{16}\left(\frac{3}{\pi N\ell^2} \right)^{3/2} \text{erf}\left(h\sqrt{\frac{3}{N\ell^2}} \right) \exp\left[-\frac{3r_{ij}^2}{4N\ell^2} \right].$$
$$\text{(A30)}$$

From these formulas, using Eq. (A5), we obtain analytical estimates of $\Delta G_{ij}^{(\text{cnf})}$. In Figure 9, these estimates are shown in the solid curve. We see an overall agreement with the exact result up to $r_{ij} = 50\,\text{nm}$, where the finite extensibility of the chain becomes important.

The above formulas can be used to derive an approximate expression for the coefficients K_{ab} used in the mean-field treatment of Section II C (Eq. (26)). Indeed, for bridges,

$$K_{ab} \approx \int_h^\infty \mathrm{d}r_{ab}\, 2\pi r_{ab} \frac{Q_{ab}^{(\text{bridge})}}{Q_a^{(\text{cnf})} Q_b^{(\text{cnf})}}, \quad \text{(A31)}$$

and for loops,

$$K_{ab} \approx \int_0^\infty \mathrm{d}r_{ab}\, 2\pi r_{ab} \frac{Q_{ab}^{(\text{loop})}}{Q_a^{(\text{cnf})} Q_b^{(\text{cnf})}}. \quad \text{(A32)}$$

In both of these cases, we have assumed that $Q_{ab}^{(\text{cnf})}$ goes to zero sufficiently quickly for the integrals to be insensitive to their upper bounds.

Finally, the above formulas also yield simple approximate expressions for the free energy of repulsion, $F_{\text{rep},i}$, of a strand i. From Eq. (A6), we obtain

$$F_{\text{rep},i} \approx -k_{\text{B}}T \ln \text{erf}\left[\sqrt{\frac{3h^2}{2N\ell^2}} \right]. \quad \text{(A33)}$$

Appendix B: Relation to Local Chemical Equilibrium treatment

The derivation of Eqs. (6) and (7) can be regarded as a "ligand-centric" result: the focus is solely on which

ligands (DNA strands) bind to one another. We can take a complementary "sticky-end-centric" approach to obtain the same results, which sheds light on the recent Local Chemical Equilibrium treatment of Rogers and Crocker[17] and corrects its main deficiency.

As in the main text, we first treat all strands separately, then make continuum approximations for simplified geometries. Following the notation of Ref. 17, let $C_i^0(\mathbf{r})$ be the normalized probability density of finding the sticky end of tether i at $\mathbf{r}$, before any hybridization occurs. For this treatment only, we assume that the strands interact "ideally", i.e., with an interaction energy ΔG_{ij}^0 when their sticky ends coincide and none otherwise, so that the configurations of the tether of i with sticky end at $\mathbf{r}$ have the same statistical weight whether the tethers are bound or unbound. In terms of partition functions, strands i and j interact ideally if their partition function when bound, Q_{ij}, is given by

$$Q_{ij} = \int d\mathbf{r} \, Q_i(\mathbf{r}) Q_j(\mathbf{r}) \frac{\exp(-\beta \Delta G_{ij}^0)}{\rho_0}, \quad \text{(B1)}$$

where $Q_i(\mathbf{r})$ is the partition function of unbound strand i with sticky end at $\mathbf{r}$, and $Q_j(\mathbf{r})$ is defined similarly. The restriction of ideal binding, absent in the general treatment of the main text, is satisfied when the tethers are ideal chains, for example. With this additional assumption, we rewrite Eq. (1) as follows:

$$\frac{Z}{Z_0} = \sum_\phi \prod_{(i,j)\in\phi} \int d\mathbf{r} \, C_i^0(\mathbf{r}) C_j^0(\mathbf{r}) \frac{\exp(-\beta \Delta G_{ij}^0)}{\rho_0}, \quad \text{(B2)}$$

where ρ_0 is the standard concentration, $1\,\text{M}$. The analog of Eq. (2) is:

$$1 = \int d\mathbf{r} \left\{ \frac{Z_{-i}}{Z} C_i^0(\mathbf{r}) \right.$$
$$\left. + \sum_j C_i^0(\mathbf{r}) C_j^0(\mathbf{r}) \frac{\exp(-\beta \Delta G_{ij}^0)}{\rho_0} \frac{Z_{-i,-j}}{Z} \right\}. \quad \text{(B3)}$$

As in the main text, let $p_i = Z_{-i}/Z$, and approximate $Z_{-i,-j}/Z$ by $p_i p_j$. As in Ref. 17, let $C_i(\mathbf{r})$ be the concentration of free sticky end i at $\mathbf{r}$, equal to $p_i C_i^0(\mathbf{r})$ by comparison to the equation above. There being only one such sticky end, $C_i(\mathbf{r})$ is also the probability density that the sticky end of tether i is free and is at $\mathbf{r}$. We obtain

$$1 = \int d\mathbf{r} \left\{ C_i(\mathbf{r}) + \sum_j C_i(\mathbf{r}) C_j(\mathbf{r}) \frac{\exp(-\beta \Delta G_{ij}^0)}{\rho_0} \right\}. \quad \text{(B4)}$$

More concisely, let $C_{ij}(\mathbf{r})$ be the concentration of bound sticky ends i and j, given by

$$C_{ij}(\mathbf{r}) = C_i(\mathbf{r}) C_j(\mathbf{r}) \frac{\exp(-\beta \Delta G_{ij}^0)}{\rho_0}. \quad \text{(B5)}$$

We can then finally write the self-consistent condition for the quantities $\{p_i\}$ as follows:

$$1 = \int d\mathbf{r} \left\{ C_i(\mathbf{r}) + \sum_j C_{ij}(\mathbf{r}) \right\}, \quad \text{(B6a)}$$

$$C_i(\mathbf{r}) = p_i C_i^0(\mathbf{r}). \quad \text{(B6b)}$$

Comparing the above condition to the equations in Ref. 17, we conclude that the "local chemical equilibrium" approximation introduced by the authors is equivalent to our approximation $Z_{-i,-j}/Z \approx p_i p_j$, but that the self-consistent condition is significantly different from the condition that they proposed,

$$C_i(\mathbf{r}) \overset{?}{=} C_i^0(\mathbf{r}) - \sum_j C_{ij}(\mathbf{r}). \quad \text{(B7)}$$

Physically, in the treatment of Rogers and Crocker, depletion of free sticky ends i at a point $\mathbf{r}$ establishes a density gradient in $C_i(\mathbf{r})$ beyond that implied by the polymer statistics of tether i, but this density gradient is not allowed to relax, effectively imposing a nonequilibrium condition. Our treatment, summarized in Eq. (B6), corrects this shortcoming.

As in the main text, for those cases where there is translational invariance, we may approximately treat together tethers with similar binding behavior. In the concrete case of two plates perpendicular to the z axis, and in the absence of binding, the concentration of a-type sticky ends at a height z is:

$$C_a^0(z) = \sigma_a P_a(z), \quad \text{(B8)}$$

where

$$P_a(z) = \int dx \, dy \, C_i^0(x, y, z), \quad \text{(B9)}$$

and i is any representative tether of type a. The corresponding mean field self-consistent equations are

$$\sigma_a = \int dz \left\{ C_a(z) + \sum_b C_{ab}(z) \right\}, \quad \text{(B10)}$$

$$C_a(z) = p_a C_a^0(z), \quad \text{(B11)}$$

$$C_{ab}(z) = C_a(z) C_b(z) \frac{\exp(-\beta \Delta G_{ab}^0)}{\rho_0}. \quad \text{(B12)}$$

Rearranging these equations to obtain explicit equations for $\{p_a\}$ yields

$$p_a = \frac{1}{1 + p_b \sigma_b \int dz \, P_a(z) P_b(z) \frac{\exp(-\beta \Delta G_{ab}^0)}{\rho_0}} \quad \text{(B13)}$$

Comparing this last equation to Eq. (21), we find that

$$K_{ab} = \frac{e^{-\beta \Delta G_{ab}^0}}{\rho_0} \int dz \, P_a(z) P_b(z). \quad \text{(B14)}$$

This is a specialization of Eq. (22) under the additional assumption of ideal binding interactions.

For the concrete case of short, rigid-rod-like tethers of length L, the probabilities $P_a(z)$ are particularly simple, and depend only on rod length, plate separation, and the plate on which a-type tethers are grafted. In particular, if they are grafted on the plate at $z = 0$ and the other plate is at $z = h$, then

$$P_a(z) = \begin{cases} 1/h, & 0 < h < L; \\ 1/L, & h > L \text{ and } 0 < z < L; \\ 0, & \text{otherwise.} \end{cases} \quad \text{(B15)}$$

Otherwise, if a-type tethers are grafted on the $z = h$ plate, then

$$P_a(z) = \begin{cases} 1/h, & 0 < h < L; \\ 1/L, & h > L \text{ and } h - L < z < L; \\ 0, & \text{otherwise.} \end{cases} \quad \text{(B16)}$$

From these results, one can easily obtain Eq. (25). This derivation should be compared with the explicit integrals involved in calculating K_{ab} using Eq. (22) and the relevant expressions for $\Delta G_{ij}^{(\text{cnf})}$ given in Appendix A of this paper and Appendix A of Ref. 16. We have verified that both routes give identical results when evaluated numerically for $L_b = L_a$ and $L_b = 1.7 L_a$.

In the absence of translational invariance or some similar symmetry, it is nontrivial, in general, to write down self-consistency conditions similar to Eq. (B6) where like-type tethers are grouped together. Unlike the incorrect condition of Eq. (B7), the correct self-consistency conditions are inherently non-local from the point of view of sticky ends: the change upon binding in sticky end concentration of a-type sticky ends at $\mathbf{r}$ depends on the *precise* identity of the a-type tether to which that sticky end is attached. Indeed, one needs to retain the level of detail inherent in Eq. (B6). As such, there is no advantage in adopting a "sticky-end-centric" point of view as opposed to a "tether-centric" point of view when dealing with arbitrarily shaped colloids, polymer statistics or non-uniform coverage. Moreover, the tether-centric equations are simpler, since they result from integrating over the irrelevant degrees of freedom describing the position of the sticky ends, and are more general, since they do not assume that binding is ideal.

To conclude this section, we present a numerical comparison between Monte Carlo results, our self-consistent theory and the LCE theory that illustrates the extent to which these approximate theories deviate from an exact treatment. We consider two periodic parallel plates, each with 500 tethers, at a density comparable to that used in Ref. 17 $(4500/(4\pi \times 550^2) = 0.001183\,\text{tethers/nm}^2)$. The tethers are modeled as ideal 8-segment chains of Kuhn length 5 nm that, when bound, form a 16-segment ideal chain, as described in Appendix A. Figure 10 shows the number of bonds formed in this system as a function of the solution hybridization free energy, ΔG^0, of the sticky
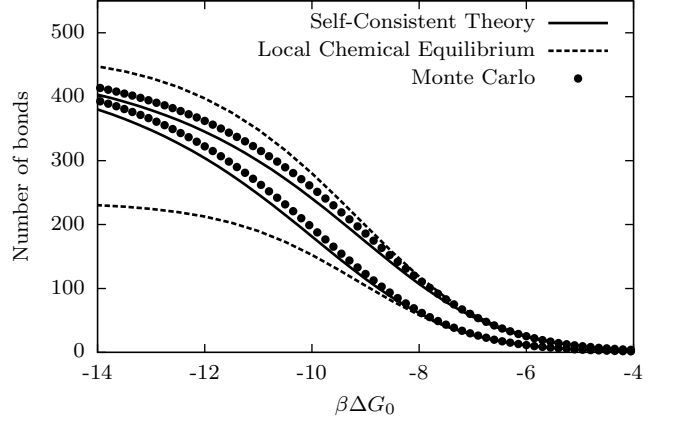

FIG. 10. Exact and estimated number of bonds formed between two ssDNA-coated parallel plates separated by 15 nm (lower curves) and 25 nm (upper curves). The Self-Consistent Theory is that of Eqs. (6) and (7), while the Local Chemical Equilibrium curve results from the methods of Ref. 17. See text for details.

ends of tethers on opposite plates, when the plates are separated by 15 nm and by 25 nm. As is observed, our self-consistent theory closely tracks the Monte Carlo results, whereas the estimates obtained using the LCE approach of Rogers and Crocker differ appreciably from the exact ones.

Appendix C: Self-consistent equations from a Saddle Point Approximation

In this appendix, we present an alternate derivation for the self-consistent mean field theory in Eqs. (21) and (22) in two specific setups, based on a saddle point approximation of a mean-field partition function. The two geometries we explore are the first two examples in Section III, which we refer to as the single linkage system and the competing linkages system.

1. Single linkage system

In the single linkage case, two parallel plates of area A are respectively decorated with N_a and N_b strands of two complementary kinds of DNA, a and b, whose surface concentration is thus $\sigma_a = N_a/A$ and $\sigma_b = N_b/A$. Without loss of generality, we assume that $N_a \leq N_b$. Let $n_{b \leftarrow a}$ be the average number of b-type strands that a fixed a-type strand can bind to. When no bonds have yet formed $n_{b \leftarrow a} = \sigma_b A_{b \leftarrow a}$, where $A_{b \leftarrow a}$ is the area that encloses all possible grafting points of b-type binding partners of a fixed a-type strand. However, after having formed n_{ab} bonds between a- and b-type strands

the average number of available linkers is reduced to[16]

$$n_{b \leftarrow a}(n_{ab}) = \frac{N_b - n_{ab}}{N_b} \sigma_b A_{b \leftarrow a} \, . \qquad \text{(C1)}$$

The average Boltzmann factor for binding a- and b-type strands is $K_{ab}/A_{b \leftarrow a}$, where K_{ab} is defined in Eq. (22). With this results and the previous one, we can write a mean-field partition function of the system as follows

$$
\begin{aligned}
Z &= \sum_{n_{ab}=0}^{N_a} \binom{N_a}{n_{ab}} \left[\prod_{j=0}^{n_{ab}-1} \frac{K_{ab}}{A_{b \leftarrow a}} n_{b \leftarrow a}(j) \right], \\
&= \sum_{n_{ab}=0}^{N_a} \binom{N_a}{n_{ab}} \binom{N_b}{n_{ab}} \frac{n_{ab}!}{N_b^{n_{ab}}} (K_{ab}\sigma_b)^{n_{ab}} \, . \qquad \text{(C2)}
\end{aligned}
$$

We now isolate the dependence of Z on A in order to take the limit $A \to \infty$, which we then evaluate through a saddle-point approximation. Let $s_{ab} = n_{ab}/A$, and approximate the sum in Eq. (C2) by an integral with respect to s_{ab}. Assuming the maximum of the integrand is far from the integration limits, we obtain

$$Z \approx A \int \mathrm{d}s_{ab} \, \exp[A \cdot \mathcal{S}(s_{ab})], \qquad \text{(C3)}$$

where, using Stirling's approximation,

$$
\begin{aligned}
\mathcal{S}(s) = {} & \sigma_a \ln \sigma_a + \sigma_b \ln \sigma_b - s \ln s - s \\
& - (\sigma_a - s)\ln(\sigma_a - s) - (\sigma_b - s)\ln(\sigma_b - s) \\
& + s \ln K_{ab} . \quad \text{(C4)}
\end{aligned}
$$

In the limit $A \to \infty$, the partition function will be dominated by the values of s that maximize $\mathcal{S}(s)$. Setting

$\mathrm{d}\mathcal{S}/\mathrm{d}s$ to 0 at $s = \sigma_{ab}$, we obtain the following equation for the average density of hybridized strands, σ_{ab}:

$$\sigma_{ab} = (\sigma_a - \sigma_{ab})K_{ab}(\sigma_b - \sigma_{ab}), \qquad \text{(C5)}$$

which restates Eq. (23) after identifying $\sigma_a p_a$ with $\sigma_a - \sigma_{ab}$ and likewise for $\sigma_b p_b$. Using Eq. (21) one can also rewrite Eq. (C5) as

$$\sigma_{ab} = \frac{\sigma_a K_{ab}(\sigma_b - \sigma_{ab})}{1 + K_{ab}(\sigma_b - \sigma_{ab})} \, , \qquad \text{(C6)}$$

which is the expression used in Ref. 16.

2. Competing linkages system

In this section we generalize the saddle point approach to the competing linkages system. Like in Section III, we have two parallel plates of area A, one grafted with a- and b-type strands, the other with a'- and b'-type strands. The only allowed interactions are a–a', a–b and a'–b'. An equivalent treatment applies when, as in Ref. 16, allowed interactions are a–a', a–b' and a'–b, so that loops (intra-particle bonds) are forbidden. The only differences between the two systems are the labels in the equations and the values of the K_{ab} factors below.

Using the same notation as in the previous section, we find that the average number of binding partners available to a fixed strand of each type is

$$n_{a' \leftarrow a}(n_{aa'}, n_{a'b'}) = \frac{N_{a'} - n_{aa'} - n_{a'b'}}{N_{a'}} \sigma_{a'} A_{a' \leftarrow a} \, , \qquad \text{(C7a)}$$

$$n_{b \leftarrow a}(n_{ab}) = \frac{N_b - n_{ab}}{N_b} \sigma_b A_{b \leftarrow a} \, , \qquad \text{(C7b)}$$

$$n_{b' \leftarrow a'}(n_{a'b'}) = \frac{N_{b'} - n_{a'b'}}{N_{b'}} \sigma_{b'} A_{b' \leftarrow a'} \, . \qquad \text{(C7c)}$$

Assume for simplicity that $N_a \leq N_{a'}$, $N_a \leq N_b$ and $N_{a'} \leq N_{b'}$. A nearly identical argument applies for a different relative populations of strands. A mean-field partition function for the system is

$$
\begin{aligned}
Z = {} & \sum_{n_{aa'}=0}^{N_a} \sum_{n_{ab}=0}^{N_a - n_{aa'}} \sum_{n_{a'b'}=0}^{N_{a'} - n_{aa'}} \binom{N_a}{n_{aa'}} \left[\prod_{k=0}^{n_{aa'}-1} \frac{K_{aa'}}{A_{a' \leftarrow a}} n_{a' \leftarrow a}(k, 0) \right] \\
& \cdot \binom{N_a - n_{aa'}}{n_{ab}} \left[\prod_{t=0}^{n_{ab}-1} \frac{K_{ab}}{A_{b \leftarrow a}} n_{b \leftarrow a}(t) \right] \binom{N_{a'} - n_{aa'}}{n_{a'b'}} \left[\prod_{v=0}^{n_{a'b'}-1} \frac{K_{a'b'}}{A_{b' \leftarrow a'}} n_{b' \leftarrow a'}(v) \right] \\
= {} & \sum_{n_{aa'}=0}^{N_a} \sum_{n_{ab}=0}^{N_a - n_{aa'}} \sum_{n_{a'b'}=0}^{N_{a'} - n_{aa'}} \binom{N_a}{n_{aa'}} \binom{N_{a'}}{n_{aa'}} \frac{n_{aa'}!}{N_{a'}^{n_{aa'}}} (K_{aa'}\sigma_{a'})^{n_{aa'}} \\
& \cdot \binom{N_a - n_{aa'}}{n_{ab}} \binom{N_b}{n_{ab}} \frac{n_{ab}!}{N_b^{n_{ab}}} (K_{ab}\sigma_b)^{n_{ab}} \binom{N_{a'} - n_{aa'}}{n_{a'b'}} \binom{N_{b'}}{n_{a'b'}} \frac{n_{a'b'}!}{N_{b'}^{n_{a'b'}}} (K_{a'b'}\sigma_{b'})^{n_{a'b'}} \, . \qquad \text{(C8)}
\end{aligned}
$$

As in the previous section, we isolate the dependence of Z on A to take the $A \to \infty$ limit. Let $s_{xy} = n_{xy}/A$, and approximate the sums in Eq. (C8) by integrals with respect to $s_{aa'}$, s_{ab} and $s_{a'b'}$. Assuming the maximum of the integrand is far from the integration limits, we obtain

$$Z \approx A^3 \iiint \mathrm{d}s_{aa'} \, \mathrm{d}s_{ab} \, \mathrm{d}s_{a'b'} \, \exp[A \cdot \mathcal{S}(s_{aa'}, s_{ab}, s_{a'b'})],$$
(C9)

where

$$\mathcal{S}(s_{aa'}, s_{ab}, s_{a'b'}) = \text{const.}$$
$$+ s_{aa'}(\ln \frac{K_{aa'}}{s_{aa'}} - 1) + s_{ab}(\ln \frac{K_{ab}}{s_{ab}} - 1) + s_{a'b'}(\ln \frac{K_{a'b'}}{s_{a'b'}} - 1)$$
$$- (\sigma_b - s_{ab})\ln[\sigma_b - s_{ab}] - (\sigma_{b'} - s_{a'b'})\ln[\sigma_{b'} - s_{a'b'}]$$
$$- (\sigma_a - s_{aa'} - s_{ab})\ln[\sigma_a - s_{aa'} - s_{ab}]$$
$$- (\sigma_{a'} - s_{aa'} - s_{a'b'})\ln[\sigma_{a'} - s_{aa'} - s_{a'b'}], \quad \text{(C10)}$$

where the first, constant term does not depend on s_{xy}.

The point $(\sigma_{aa'}, \sigma_{ab}, \sigma_{a'b'})$ at which $\mathcal{S}$ is maximized, found by setting all its partial derivatives to zero, yields the average bond densities for each pair of species. They satisfy

$$\sigma_{aa'} = (\sigma_a - \sigma_{aa'} - \sigma_{ab})K_{aa'}(\sigma_{a'} - \sigma_{aa'} - \sigma_{a'b'})$$
$$= \sigma_a p_a K_{aa'} p_{a'} \sigma_{a'}, \quad \text{(C11a)}$$
$$\sigma_{ab} = (\sigma_a - \sigma_{aa'} - \sigma_{ab'})K_{ab}(\sigma_b - \sigma_{ab})$$
$$= \sigma_a p_a K_{ab} p_b \sigma_b, \quad \text{(C11b)}$$
$$\sigma_{a'b'} = (\sigma_{a'} - \sigma_{aa'} - \sigma_{a'b'})K_{a'b'}(\sigma_{b'} - \sigma_{a'b'})$$
$$= \sigma_{a'} p_{a'} K_{a'b'} p_{b'} \sigma_{b'} . \quad \text{(C11c)}$$

These equations correspond to the mean-field self-consistent theory of Eqs. (21) and (23)) for the competing linkages model.

We now show that the saddle points equations (C11) are also equivalent to the expressions derived in Ref. 16 (once σ_{ab} and $\sigma_{a'b'}$ are replaced with $\sigma_{ab'}$ and $\sigma_{a'b}$ respectively). Using Eq. (C11c) two times, we have

$$\sigma_{a'} - \sigma_{aa'} - \sigma_{a'b'}$$
$$= \frac{\sigma_{a'} - \sigma_{aa'}}{(\sigma_{a'} - \sigma_{aa'} - \sigma_{a'b'}) + \sigma_{a'b'}} \cdot \frac{\sigma_{a'b'}}{K_{a'b'}(\sigma_{b'} - \sigma_{a'b'})}$$
$$= \frac{\sigma_{a'} - \sigma_{aa'}}{1 + K_{a'b'}(\sigma_{b'} - \sigma_{a'b'})} . \quad \text{(C12)}$$

Substituting this result into Eq. (C11a), together with the definition of p_a in Eq. (21), we find that

$$\sigma_{aa'} = \sigma_a K_{aa'} \frac{\sigma_{a'} - \sigma_{aa'}}{1 + K_{a'b'}(\sigma_{b'} - \sigma_{a'b'})}$$
$$\cdot \left[1 + \frac{K_{aa'}(\sigma_{a'} - \sigma_{aa'})}{1 + K_{a'b'}(\sigma_{b'} - \sigma_{a'b'})} + K_{ab}(\sigma_b - \sigma_{ab})\right]^{-1}$$
$$= \sigma_a K_{aa'}(\sigma_{a'} - \sigma_{aa'}) \cdot \left[1 + K_{a'b'}(\sigma_{b'} - \sigma_{a'b'})\right.$$
$$+ K_{aa'}(\sigma_{a'} - \sigma_{aa'}) + K_{ab}(\sigma_b - \sigma_{ab})$$
$$\left. + K_{ab}(\sigma_b - \sigma_{ab}) \cdot K_{a'b'}(\sigma_{b'} - \sigma_{a'b'})\right]^{-1} .$$
(C13)

This is exactly the expression used in Ref. 16 to compute the density of a–a' linkages, derived there by applying an heuristic self-consistent prescription to an approximate estimate of $\sigma_{aa'}$. A similar argument recovers the expressions for σ_{ab} and $\sigma_{a'b'}$ used in Ref. 16.

[1] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff, "A DNA-based method for rationally assembling nanoparticles into macroscopic materials," Nature **382**, 607–609 (1996).

[2] A. P. Alivisatos, K. P. Johnsson, X. G. Peng, T. E. Wilson, C. J. Loweth, M. P. Bruchez, and P. G. Schultz, "Organization of 'nanocrystal molecules' using DNA," Nature **382**, 609–611 (1996).

[3] P. L. Biancaniello, A. J. Kim, and J. C. Crocker, "Colloidal interactions and self-assembly using DNA hybridization," Phys. Rev. Lett. **94**, 58302 (2005).

[4] N. Geerts and E. Eiser, "DNA-functionalized colloids: Physical properties and applications," Soft Matter **6**, 4647–4660 (2010).

[5] N. L. Rosi and C. A. Mirkin, "Nanostructures in biodiagnostics," Chem. Rev. **105**, 1547–1562 (2005).

[6] D. Nykypanchuk, M. M. Maye, D. van der Lelie, and O. Gang, "DNA-guided crystallization of colloidal nanoparticles," Nature **451**, 549–552 (2008).

[7] S. Y. Park, A. K. R. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin, "DNA-programmable nanoparticle crystallization," Nature **451**, 553–556 (2008).

[8] M. M. Maye, M. T. Kumara, D. Nykypanchuk, W. B. Sherman, and O. Gang, "Switching binary states of nanoparticle superlattices and dimer clusters by DNA strands," Nat. Nanotechnol. **5**, 116–120 (2010).

[9] R. J. Macfarlane, B. Lee, M. R. Jones, N. Harris, G. C. Schatz, and C. A. Mirkin, "Nanoparticle superlattice engineering with DNA." Science **334**, 204–8 (2011).

[10] N. A. Licata and A. V. Tkachenko, "Statistical mechanics of DNA-mediated colloidal aggregation," Phys. Rev. E **74**, 41408 (2006).

[11] D. Nykypanchuk, M. M. Maye, D. van der Lelie, and O. Gang, "DNA-based approach for interparticle interaction control." Langmuir **23**, 6305–14 (2007).

[12] R. Dreyfus, M. E. Leunissen, R. Sha, A. V. Tkachenko, N. C. Seeman, D. J. Pine, and P. M. Chaikin, "Simple Quantitative Model for the Reversible Association of DNA Coated Colloids," Phys. Rev. Lett. **102**, 48301 (2009).

[13] R. Dreyfus, M. E. Leunissen, R. Sha, A. V. Tkachenko, N. C. Seeman, D. J. Pine, and P. M. Chaikin, "Aggregation-disaggregation transition of DNA-coated colloids: Experiments and theory," Phys. Rev. E **81**, 41404 (2010).

[14] M. E. Leunissen, R. Dreyfus, R. Sha, N. C. Seeman, and P. M. Chaikin, "Quantitative Study of the Association Thermodynamics and Kinetics of DNA-Coated Particles for Different Functionalization Schemes," J. Am. Chem. Soc. **132**, 1903–1913 (2010).

[15] M. E. Leunissen and D. Frenkel, "Numerical study of DNA-functionalized microparticles and nanoparticles: Explicit pair potentials and their implications for phase behavior," J. Chem. Phys. **134**, 84702 (2011).

[16] B. M. Mognetti, M. E. Leunissen, and D. Frenkel, "Controlling the temperature sensitivity of DNA-mediated colloidal interactions through competing linkages," Soft Matter **8**, 2213–2221 (2012).

[17] W. B. Rogers and J. C. Crocker, "Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling," P. Natl. Acad. Sci. U.S.A. **108**, 15687–15692 (2011).

[18] B. M. Mognetti, P. Varilly, S. Angioletti-Uberti, F. J. Martinez-Veracoechea, J. Dobnikar, M. E. Leunissen, and D. Frenkel, "Predicting DNA-mediated colloidal pair interactions," P. Natl. Acad. Sci. U.S.A. **109**, E378–E379 (2012).

[19] W. B. Rogers and J. C. Crocker, "Reply to Mognetti et al.: DNA handshaking interaction data are well described by mean-field and molecular models," P. Natl. Acad. Sci. U.S.A. **109**, E380–E380 (2012).

[20] E. Zaccarelli, I. Saika-Voivod, S. V. Buldyrev, A. J. Moreno, P. Tartaglia, and F. Sciortino, "Gel to glass transition in simulation of a valence-limited colloidal system." J. Chem. Phys. **124**, 124908 (2006).

[21] F. J. Martinez-Veracoechea and D. Frenkel, "Designing super selectivity in multivalent nano-particle binding," P. Natl. Acad. Sci. U.S.A. **108**, 10963–10968 (2011).

[22] V. M. Krishnamurthy, L. A. Estroff, and G. M. Whitesides, "Multivalency in Ligand Design," in *Fragment-based Approaches in Drug Discovery*, Methods and Principles in Medicinal Chemistry, edited by W. Jahnke and D. A. Erlanson (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, FRG, 2006) Chap. 2, pp. 11—53.

[23] J. D. Badjić, A. Nelson, S. J. Cantrill, W. B. Turnbull, and J. F. Stoddart, "Multivalency and cooperativity in supramolecular chemistry." Accounts Chem. Res. **38**, 723–32 (2005).

[24] M. Mammen, S.-K. Choi, and G. M. Whitesides, "Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors," Angew. Chem. Int. Edit. **37**, 2754–2794 (1998).

[25] S. Angioletti-Uberti, B. M. Mognetti, and D. Frenkel, "Reentrant melting as a design principle for DNA-coated colloids," Nat. Mater. **11**, 518–522 (2012).

[26] Although one can design a strand that binds more than one partner simultaneously, doing so is uncommon for DNACCs, since the entropic cost of bringing three or more strands together when they are grafted onto colloids (as opposed to in solution) is prohibitive.

[27] M. P. Valignat, O. Theodoly, J. C. Crocker, W. B. Russel, and P. M. Chaikin, "Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids," P. Natl. Acad. Sci. U.S.A. **102**, 4225–4229 (2005).

[28] Z. Tan, E. Gallicchio, M. Lapelosa, and R. M. Levy, "Theory of binless multi-state free energy estimation with applications to protein-ligand binding," J. Chem. Phys. **136**, 144102 (2012).

[29] F. Zhu and G. Hummer, "Convergence and error estimation in free energy calculations using the weighted histogram analysis method," J. Comput. Chem. **33**, 453–465 (2012).

[30] O. Taussky, "A Recurring Theorem on Determinants," Am. Math. Mon. **56**, 672—676 (1949).

[31] J. SantaLucia Jr. and D. Hicks, "The thermodynamics of DNA structural motifs," Annu. Rev. Biophys. Bio. **33**, 415–440 (2004).

[32] R. J. Hunter, *Foundations of Colloid Science* (Oxford University Press, 2000).

[33] Self-contained programs to calculate the SCT and SCTMF predictions for all of these examples are included in the accompanying code, in the `examples` folder.

[34] N. Kern and D. Frenkel, "Fluidfluid coexistence in colloidal systems with short-ranged strongly directional attraction," J. Chem. Phys. **118**, 9882 (2003).

[35] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," Nature **440**, 297–302 (2006).

[36] Q. Chen, S. C. Bae, and S. Granick, "Directed self-assembly of a colloidal kagome lattice," Nature **469**, 381–384 (2011).

[37] An ambiguity arises from including or excluding the bases of the tether immediately adjacent to the sticky end, since they can form stacking interactions with the dsDNA formed by two bound sticky ends. If included, DINAmelt predicts $\Delta H^0 = -56.7 \, \text{kcal/mol}$ and $\Delta S^0 = -166.0 \, \text{cal/(mol·K)}$. Since the scope of this paper is to evaluate the accuracy of the SCT treatment with respect to full MC simulations, we sidestep any attempt to resolve this ambiguity.

[38] N. R. Markham and M. Zuker, "DINAMelt web server for nucleic acid melting prediction." Nucleic Acids Res. **33**, W577–81 (2005).

[39] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications (Computational Science)* (Academic Press Inc, 2001).

[40] S. Torquato, "Inverse optimization techniques for targeted self-assembly," Soft Matter **5**, 1157 (2009).

[41] G. I. Bell, M. Dembo, and P. Bongrand, "Cell adhesion. Competition between nonspecific repulsion and specific bonding." Biophys. J. **45**, 1051–64 (1984).

[42] T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, "Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model." J. Chem. Phys. **134**, 085101 (2011).

[43] One can imagine coalescing more segments than the final one. However, since the bound site is presumably poorly modeled by a FJC of the same persistence length as the unbound polymers, one would have to go beyond the treatment given here.

[44] L. R. G. Treloar, "The statistical length of long-chain molecules," T. Faraday Soc. **42**, 77 (1946).

[45] H. Yamakawa, *Modern Theory of Polymer Solutions*, electronic ed. (Harper, 1971).

[46] J. S. Liu, *Monte Carlo Strategies in Scientific Computing (Springer Series in Statistics)* (Springer, 2002).